DOCUMENT RESUME

ED 320 945 TM 015 215

AUTHOR Schmidt, Henk G.; And Others
TITLE Use of Student Ratings for Program Improvement: Examples from a Problem-Based Medical Curriculum.
PUB DATE Apr 90
NOTE 6p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990). For a related document, see TM 015 216.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Curriculum Development; *Curriculum Evaluation; *Evaluation Utilization; Foreign Countries; Higher Education; Likert Scales; Medical Schools; *Medical Students; Program Evaluation; *Program Improvement; *Rating Scales; *Student Evaluation of Teacher Performance; Test Use
IDENTIFIERS *Problem Based Medical Curriculum; Quebec

ABSTRACT

The ways in which a rating scale developed for problem-based medical curricula provides information that is useful in the area of curriculum reform were demonstrated in a study involving 95 first-year medical students at the University of Sherbrooke in Quebec (Canada). The rating scale, based on the theory of problem-based learning of W. H. Gijselaers and H. G. Schmidt (1989), covers educationally important dimensions through 10 subscales, most with Likert items. A new problem-based, community-oriented curriculum was instituted at the university in 1987. After each curriculum unit, the scale was administered to all students. Results were used to evaluate three variables: (1) quality of problems used; (2) adequacy of lectures; and (3) tutor performance. The problem of the absence of absolute standards for what can be considered sufficient performance of a course is circumvented by applying a reliable measurement approach in which each course or teacher is compared to other courses and teachers, and in which the highest score constitutes the norm against which others are judged. Four graphs illustrate the evaluation. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED320945

Use of Student Ratings for Program Improvement: Examples from a Problem-Based Medical Curriculum

Henk G. Schmidt, Jacques E. Des Marchais[o] and Robert Black[o]
University of Limburg, The Netherlands and
[o]Université de Sherbrooke, Québec, Canada

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

HENK G. SCHMIDT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Paper presented at the Meeting of the American Educational Research Association, Boston, MA., 1990

TM015215

Student ratings are generally considered reliable and valid indicators of the quality of instruction, in particular when students are asked to judge elements of the instructional context which are readily observable, like teaching skill or the adequacy of instructional materials (Cohen, 1982). Two problems, however, limit their usefulness in everyday instructional development. The first is that student ratings are often insufficiently specific to provide guidelines for course *improvement* . For instance, knowing that his lectures were judged rather unfavorably doesn't help a teacher in finding ways to improve on their quality. A way to deal with this problem is to formulate descriptive statements derived from theoretical notions of what constitute important facets of the on-going teaching and learning processes. It is, for instance, more useful to inquire about the extent to which the subject-matter studied was adapted to the level of the students' prior knowledge than to ask whether they liked the program, because amount of prior knowledge influences the processing of new information (Anderson, 1977), whereas like or dislike of instruction has no known influence on learning. The second problem is, that, since no absolute standards exist for sufficient instructional quality, it is almost impossible to decide when remedial action is required with respect to a certain course. Therefore, an approach to course improvement will be illustrated here, that is based on *comparisons* among courses.

In this paper, a study will be reported, demonstrating ways in which a rating scale developed for problem-based medical curricula provides information useful for curriculum reform. The rating scale covers educationally important dimensions and is based on a theory of problem-based learning[1] (Gijselaers & Schmidt, 1989). It consists of ten subscales, each of which has sufficient internal consistency, interrater agreement and construct validity (Des Marchais, Schmidt & Black, 1988; Gijselaers, 1988). The subscales are briefly described in Table 1. Most of the items are of the Likert type; they are to be rated on a 5-point scale ranging from "highly disagree" to "highly agree."

Table 1. Sample items descriptive for problem-based learning (Des Marchais, Schmidt & Black, 1988)

1. Tutor as contributor to learning	**7. Breadth of learning**
33. His way of interrupting disturbed the progress of the group discussion	14. In the course of the unit I have learnt other things not related to the problems themselves
26. The tutor's questions stimulated the discussion	13. The problems helped me in integrating the basic with the clinical sciences
2. Satisfaction with unit	**8. Amount of prior knowledge/Difficulty level**
6. The unit was well-organized	3. The unit's subject-matter was difficult to understand
7. The unit elapsed in a harmonious manner	2. The unit's subject-matter was adapted to my prior knowledge
3. Group functioning	**9. Time-on-task**
22. I found the atmosphere in my group agreeable	43. How much time on the average did you spent each week on independent study? (Fill in the answer in whole hours)
21. The meetings stimulated self-directed learning activities	**10. Systematic approach to learning**
4. Quality of problems	16. My tutorial group systematically applied the steps involved in problem-based learning
8. The problems were clearly stated	
12. The problems sufficiently stimulated self-directed learning	
5. Adequacy of lectures	
40. The lectures related well to the subjects I studied	
36. The lectures fitted to the theme of the unit	
6. Tutor as personal guide	
30. The tutor has helped me personally to progress	
29. The tutor has given me feedback during formative evaluation	

METHOD

Subjects and Curriculum. Subjects were 95 first-year students of the Faculté de Médecine of the Université de Sherbrooke, Québec, Canada. This medical school is involved in a large-scale transition process in which a four-year discipline-based conventional curriculum is gradually being changed to a problem-based, community-oriented program. Therefore, careful program monitoring is considered an essential element of the innovation. The first group of students enrolling in the renewed curriculum started September 1987 and

[1]Problem-based learning can be characterized as follows: A collection of carefully constructed problems is presented to small groups of students. In medical education, they usually take the form of a description of a patient, presenting with a complaint and with a number of signs and symptoms. The task of the group is to discuss these problems and produce tentative explanations for the phenomena described in terms of some underlying process, principle or mechanism. Essential to the method is, that students' prior knowledge of the problem is, in itself, insufficient to understand it in depth, so that during initial analysis dilemma's will arise and questions will come up that can be used as learning goals for subsequent, individual, self-directed learning. While analyzing a problem in a prescribed, systematic fashion, the group is guided by a tutor. His or her task is to stimulate the discussion whenever necessary, to provide students with subject-matter information when adequate and to evaluate progress being made. References, audiovisual aids and occasional lectures are included as learning resources relevant to the understanding of the problems.

has now completed seven six-week courses or "units" which comprise the first curriculum year. Table 2 lists the seven units.

Table 2. Contents of the first curriculum year, Faculté de Médecine, Université de Sherbrooke

Unit 1: Medical biology 1
Unit 2: Medical biology 2
Unit 3: Growth, development and aging
Unit 4: Nervous system
Unit 5: Locomotor system
Unit 6: Psychiatric problems
Unit 7: Community health

Procedure. After each unit, the rating scale was administered to all students. The average response rate was 86%. The total number of rating scales returned was 566. Since the data of interest were comparisons between the units, the ratings were averaged over students and subscales. Hence, the results could be interpreted as scores on the same five-point Likert scale used for the individual items. In the comparisons among units, only differences larger than .5 were considered (Differences as small as .10 are statistically significant in many cases due to the large number of subjects involved in the investigation and the high interrater agreements. Thus, statistical significance is of little use as a criterion of what constitutes a *meaningful* difference between the scores of two units).

RESULTS and DISCUSSION

Results of the evaluation of Sherbrooke's new first-year curriculum will be illustrated with regard to three variables: Quality of problems used, adequacy of the lectures and tutor performance. It will be shown how comparative data can be used for spotting areas of weakness. In addition, strategies for remediation will be briefly discussed.

Quality of problems. As displayed in Figure 1, the units of the faculty's new first-year curriculum show fairly large differences with respect to the problems used as a stimulus for learning. Ratings vary between neutral and high. Since the average scores signify a response to a statement like "The quality of the problems used was sufficient," the rule here is that problems should be rated as high as possible. Consequently, the data suggest that in at least four units improvements regarding the nature of the problems may be necessary. The question of course, is: Which improvements? A general overview like the one presented in Figure 1 does not provide answers to this question. It points at where weaknesses in particular units may reside but does not, in itself, provide suggestions for remediation. In this case, it may be useful to analyze the response patterns on individual items comprising the dimension of interest.

Figure 1. Quality of the problems used as a stimulus for learning

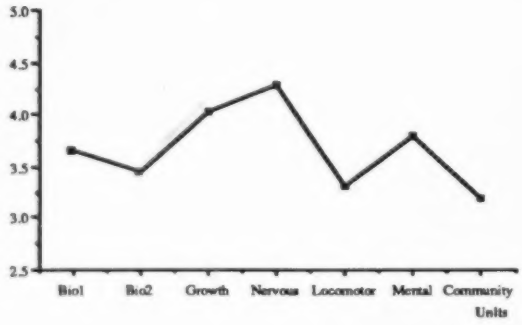
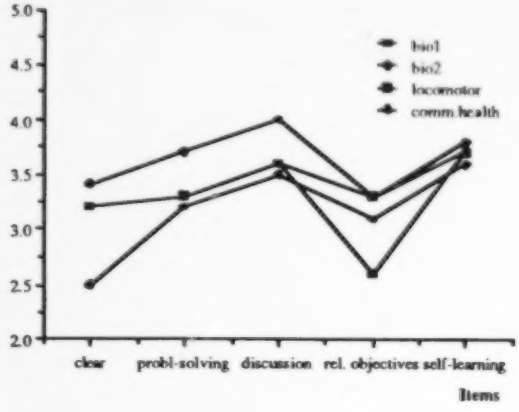


Figure 2 shows average scores on the following items: " The problems were clearly stated", "The problems were suitable for using a systematic (problem-solving) approach", "The problems sufficiently stimulated group discussion", "The problems were sufficiently related to the unit's objectives" and: "The problems sufficiently stimulated self-directed learning." It suggests that the problems used in different units may not suffer similar weaknesses. For instance, the community-health unit's problems lack sufficient clarity, as compared with problems in the other units, whereas the problems in the locomotor unit were insufficiently related to the unit's objectives. At this point, several strategies for improvement are possible. The first is to look into the problems themselves from the perspectives provided by the ratings. Often, the problems' shortcomings and possible ways by which they could be reformulated present themselves easily. A second strategy would be to interview students and staff, because sometimes, the difficulty does not so much lay in the problem formulation but in the instructional context within which the problems had to be interpreted. For instance, according to students, their tutors in the locomotor unit voiced ideas about the unit's objectives that could not be deduced from the problems themselves.

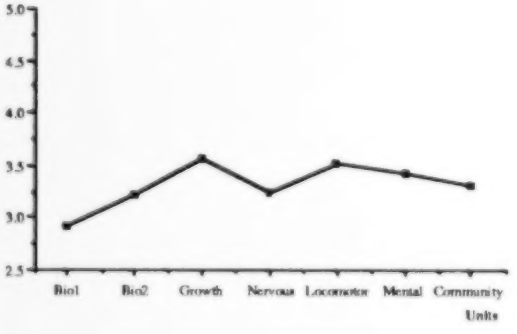
Figure 2. Average scores on five items constituting the quality-of-problems scale (Items are displayed on the horizontal axis)



Adequacy of lectures. The adequacy of the lecturing shows a pattern quite different from that of the problems. The range is about the same and differences are not large. Since one would like the lectures to be judged as adequate as possible, Figure 3 definitely points to a problem within the curriculum, in particular within the first unit on biomedical problems. More detailed analyses of the response patterns on individual items reveals that the lectures did not relate very well to the subjects studied by the students (The item pertaining to this issue was consistently rated lowest). This problem appears not to be specific to the Sherbrooke curriculum. In fact, it can be observed in other problem-based curricula as well. For instance, the average rating on the same item provided by students in the problem-based health sciences curriculum at the University of

Limburg was 3.4, suggesting that students don't clearly perceive a relation between the contents of the lectures and their own learning activities. This difficulty originates from the fact that, within the context of a problem-based curriculum, students' learning activities are always to some extent unpredictable. They are dependent on the learning goals formulated as a result of the small-group analysis of the problems. Since these learning goals tend to vary somewhat between different groups, it is almost impossible to fit the content of the lectures to the learning needs of the students. That is why many proponents of problem-based learning are reluctant to introduce a great number of lectures as part of a curriculum: Not so much because traditional lectures represent an instructional philosophy incompatible with problem-based learning, but because they tend to be ineffective. A solution sometimes proposed is to tune lectures to the questions students have regarding subject-matter

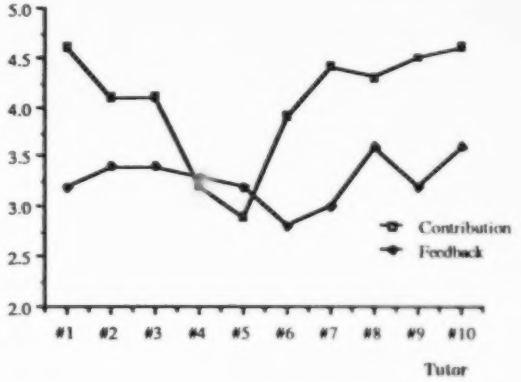
Figure 3. Adequacy of the lectures given



they are studying. To do that however, requires considerable skill from the instructor, because, since students decide on the agenda for such a meeting, it is difficult to prepare.

Tutor performance. The interpretation and use of the data provided by the rating scale is not limited to comparisons between *units* in order to arrive at conclusions regarding the quality of instruction. The data can also by used to judge the contributions of individual tutors within one single unit. As an example, average performance of the ten tutors active in the community health unit is plotted in Figure 4. They were rated on two general dimensions: How well they contributed to the learning of the group by asking stimulating questions, encourage students to work hard and intervening whenever necessary, and: The extent to which they provided students with personal feedback. Both dimensions are displayed in Figure 4. The figure shows that, according to the students, two tutors, #4 and #5, did less well in contributing to the learning process compared to their colleagues. There may be many reasons for such judgements and closer scrutiny of the responses to the various items that comprise the dimension may reveal some of them. It may be possible that these tutors were less knowledgeable about the topic of the unit and hence, were not really able to help their students in making sense out of the problems. Lack of knowledge of the subject-matter to be studied is a serious handicap for every tutor, because it makes it almost impossible, perhaps with the exception of the most experienced tutors, to sense when students are in trouble and may be helped by a clarifying question or comment. This would be in particular problematic for the community health unit, since its problems were shown to be formulated relatively unclear (See Figure 2).

Figure 4. Tutor performance in community health unit



CONCLUSION

In this paper, an approach to program evaluation is demonstrated which tries to avoid two of the pitfalls of common questionnaire-based evaluation. First, unlike other approaches, the method illustrated here is explicitly based on theoretical notions regarding the nature of the teaching-learning process taking place in problem-based curricula and on variables shown to be crucial to successful learning (Gijselaers, & Schmidt, 1989). Second, the problem of the absence of absolute standards for what can be considered sufficient performance of a course is circumvented by applying a relative-measurement approach in which each course or teacher is compared to other courses and teachers and in which the highest score constitutes the norm against which others are judged.

REFERENCES

Anderson, R. C. (1977). The notion of schemata and the educational enterprise. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.). *Schooling and the Acquisition of Knowledge*. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51,* 281-309.

Des Marchais, J. E., Schmidt, H. G., & Black, R. (1988). *Theory-guided Design of a Rating Scale for Program Evaluation in Problem-Based Medical Curricula*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Gijselaers, W. H. (1988). *Kwaliteit van Onderwijs Gemeten* (Measuring instructional quality). Ph.D.-Thesis. Maastricht, The Netherlands: University of Limburg.

Gijselaers, W. H., & Schmidt, H. G. (1989). The development and evaluation of a causal model of problem-based learning. In Z. Nooman, H. G. Schmidt, & E. Ezzat (Eds.). *Innovation in Medical Education: An Evaluation of its Present Status*. New York, NY: Springer Publishing.